

1 **CLAIMS**

2 I claim:

3
4 1. A computer-implemented method comprising:
5 defining a set of reduced regular expressions for particular patterns in
6 strings; and

7 learning, from a training set, a knowledge base that uses the reduced regular
8 expressions to resolve ambiguity based upon the strings in which the ambiguity
9 occurs, wherein the learning includes transformation sequence learning to create a
10 set of rules that use the reduced regular expressions to resolve ambiguity based
11 upon the strings in which the ambiguity occurs.

12
13 2. A computer-implemented method as recited in claim 1, wherein the
14 set of reduced regular expressions are defined over a finite alphabet Σ , wherein
15 the alphabet is a union of multiple sets of distinct classes.

16
17 3. A computer-implemented method as recited in claim 1, wherein the
18 training set comprises a labeled corpus.

19
20 4. A computer-implemented method as recited in claim 1, wherein the
21 set of reduced regular expressions specify types of patterns that are allowed to be
22 explored when learning from the training set.

1 5. A computer-implemented method as recited in claim 1, wherein the
2 learning includes applying a set of very reduced regular expressions that are a
3 proper subset of the reduced regular expressions.

4
5 6. A computer readable medium having computer-executable
6 instructions that, when executed on a processor, perform a method comprising:

7 defining a set of reduced regular expressions for particular patterns in
8 strings; and

9 learning, from a training set, a knowledge base that uses the reduced regular
10 expressions to resolve ambiguity based upon the strings in which the ambiguity
11 occurs, wherein the set of reduced regular expressions specify types of patterns
12 that are allowed to be explored when learning from the training set.

13
14 7. A computer readable medium as recited in claim 6, wherein the set
15 of reduced regular expressions are defined over a finite alphabet Σ , wherein the
16 alphabet is a union of multiple sets of distinct classes.

17
18 8. A computer-implemented method as recited in claim 6, wherein the
19 training set comprises a labeled corpus.

20
21 9. A computer-implemented method as recited in claim 6, wherein the
22 learning comprises transformation sequence learning to create a set of rules that
23 use the reduced regular expressions to resolve ambiguity based upon the strings in
24 which the ambiguity occurs.

1 10. A computer-implemented method as recited in claim 6, wherein the
2 learning includes applying a set of very reduced regular expressions that are a
3 proper subset of the reduced regular expressions.

4
5 11. A computer-implemented method comprising:
6 receiving a string with an ambiguity site;
7 applying reduced regular expressions to describe a pattern in the string,
8 wherein the reduced regular expressions:
9 are included in a knowledge base that is learned from a training set;
10 and
11 specify types of patterns that are allowed to be explored when the
12 knowledge base is learned; and
13 selecting one of the reduced regular expressions to resolve the ambiguity
14 site.

15
16 12. A computer-implemented method as recited in claim 11, wherein the
17 applying includes applying a set of very reduced regular expressions that are a
18 proper subset of the reduced regular expressions.

19
20 13. A computer-implemented method comprising:
21 receiving a string with an ambiguity site;
22 applying reduced regular expressions to describe a pattern in the string,
23 wherein the applying includes applying a set of very reduced regular expressions
24 that are a proper subset of the reduced regular expressions; and
25

1 selecting one of the reduced regular expressions to resolve the ambiguity
2 site.

3
4 14. A computer readable medium having computer-executable
5 instructions that, when executed on a processor, perform a method comprising:

6 receiving a string with an ambiguity site;

7 applying reduced regular expressions to describe a pattern in the string,

8 wherein:

9 the reduced regular expressions are included in a knowledge base
10 that is learned from a training set; and

11 the reduced regular expressions specify types of patterns that are
12 allowed to be explored when the knowledge base is learned; and

13 selecting one of the reduced regular expressions to resolve the ambiguity
14 site.

15
16 15. A computer readable medium as recited in claim 14, wherein the
17 applying includes applying a set of very reduced regular expressions that are a
18 proper subset of the reduced regular expressions.

19
20 16. A computer readable medium having computer-executable
21 instructions that, when executed, direct a computer to:

22 read a training set;

23 construct a graph having a root node that contains a primary position set of
24 the training set and multiple paths from the root node to secondary nodes that
25

1 represents a reduced regular expression, the secondary node containing a
2 secondary position set to which the reduced regular expression maps;
3 score the secondary nodes to identify a particular secondary node; and
4 identify the reduced regular expression that maps the path from the root
5 node to the particular secondary node.

6
7 17. A training system comprising:
8 a memory to store a training set;
9 a processing unit; and
10 a disambiguation trainer, executable on the processing unit, to define a set
11 of reduced regular expressions for particular patterns in strings of the training set
12 and learn a knowledge base that uses the reduced regular expressions to describe
13 the strings wherein the reduced regular expressions specify types of patterns that
14 are allowed to be explored when the knowledge base is learned from the training
15 set.

16
17 18. A training system as recited in claim 17, wherein the training set
18 comprises a labeled corpus.

19
20 19. A training system as recited in claim 17, wherein the disambiguator
21 trainer employs transformation sequence learning to create a set of rules that use
22 the reduced regular expressions to describe the strings.

23
24 20. A system comprising:
25

1 a memory to store a knowledge base that uses reduced regular expressions
2 to resolve ambiguity based upon strings in which the ambiguity occurs, wherein
3 the knowledge base is learned from a training set using the reduced regular
4 expressions, the reduced regular expressions specify types of patterns that are
5 allowed to be explored when the knowledge base is learned;

6 a processing unit; and

7 a disambiguator, executable on the processing unit, to receive a string with
8 an ambiguity site and apply a reduced regular expression from the knowledge base
9 that describes a pattern in the string to resolve the ambiguity site.